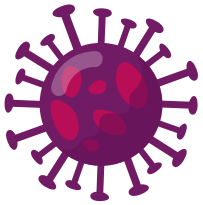


# COVID-19 Impact On Counties

**Katherine Lin**  
Byram Hills High School

Supervised by Linda Zhao

# My COVIDpath to Data Scientist



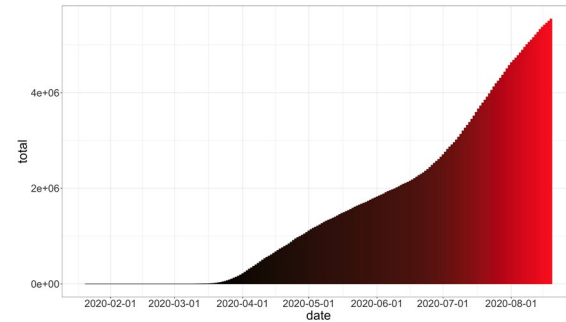
First COVIDcase

Jan

Mar

Stay at home

Soaring cases  
Despair



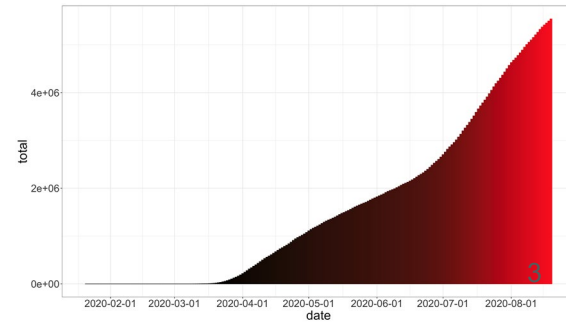
# My COVIDpath to Data Scientist

Prof Zhao's  
Modern Data Mining Course

What can I do to help?

First COVIDcase

Stay at home  
Soaring cases  
Despair



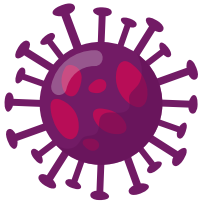


# My COVIDpath to Data Scientist

Prof Zhao's  
Modern Data Mining Course

What can I do to help?

First COVIDcase



Jan

Mar

Stay at home  
Despair

Jun

May

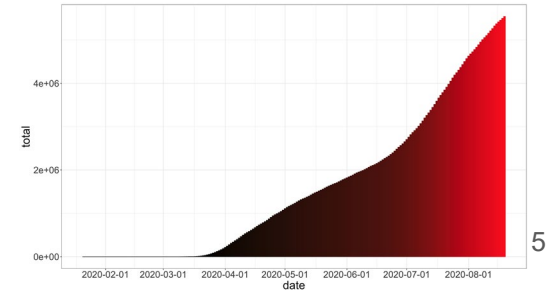
Jul

2021

Data Scientist

Aim to be a data scientist  
Use data for social good

COVID+ County study



# COVID Death Rate vs County Socioeconomic Status

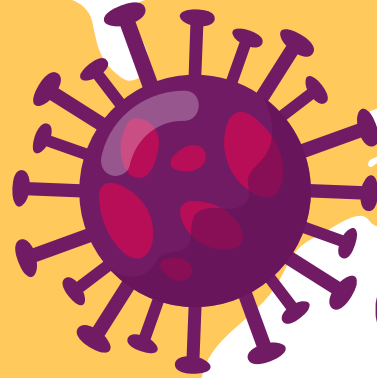
## Race group

Is one racial group affected more?



## Age group

Elderly is affected more. Is this true?



## Education level

More educated  
More informed and careful?



## Income group

High income less affected?



# COVID Death Rate vs County Socioeconomic Status

**1. Important factors related to COVID death rate**

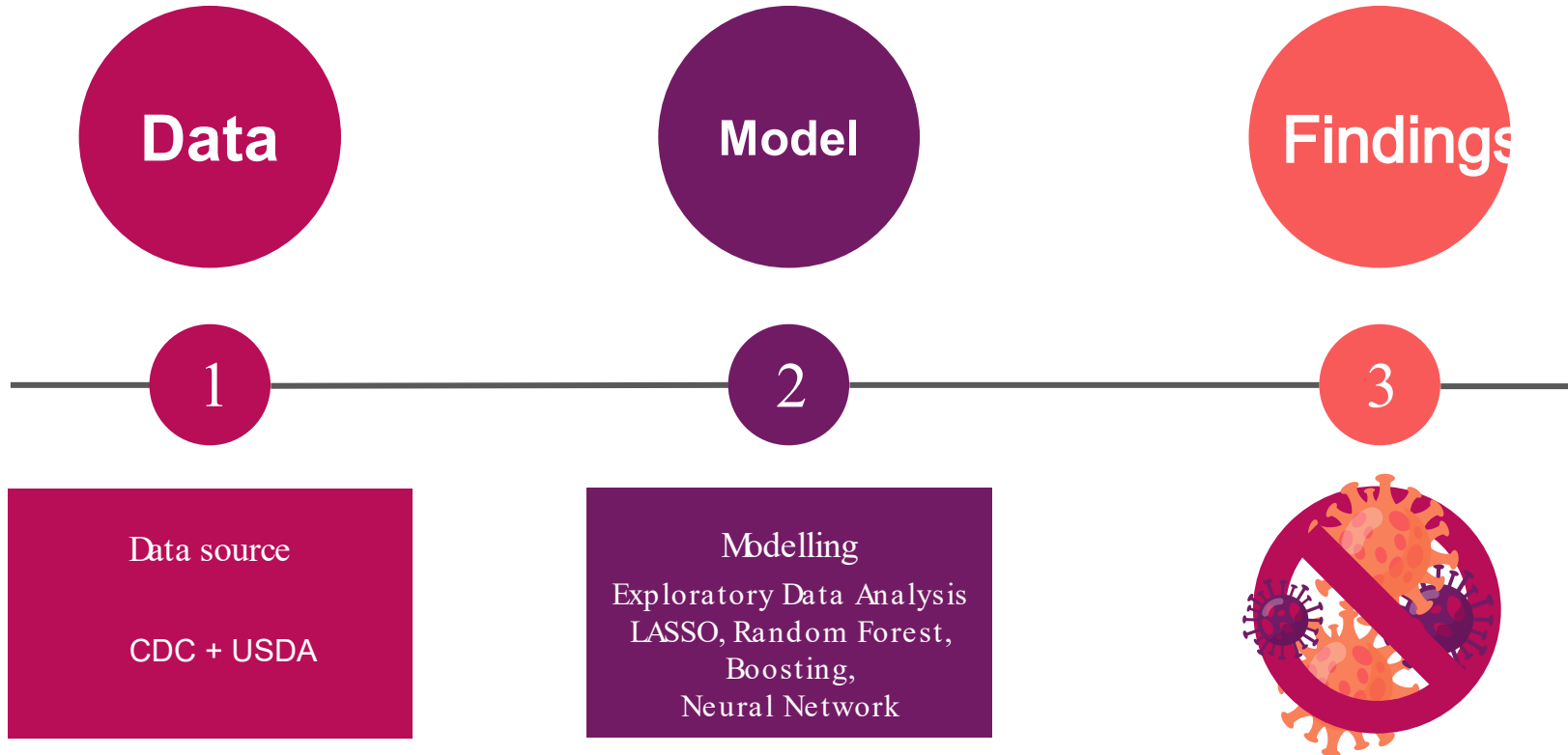
Age group

Elderly is affected more. Is this true?

**1. Predict COVID death rate**

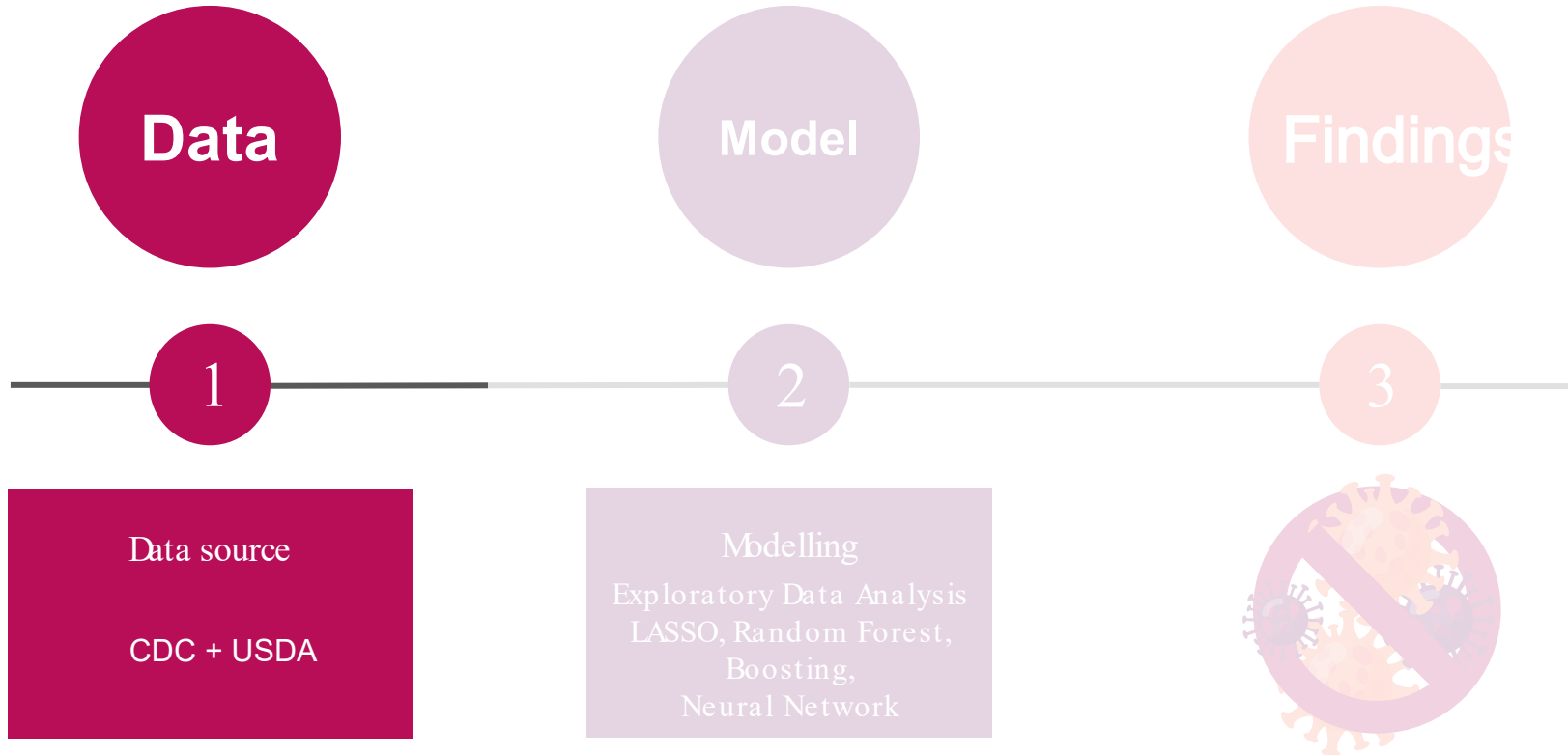
High income less affected?

# Roadmap

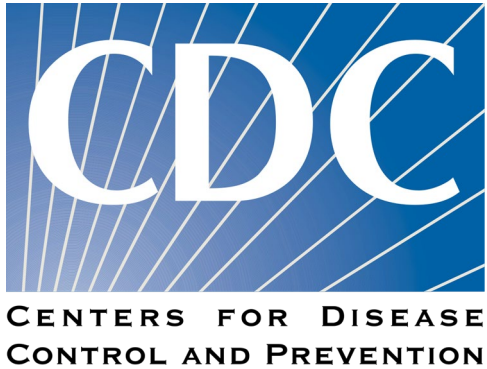




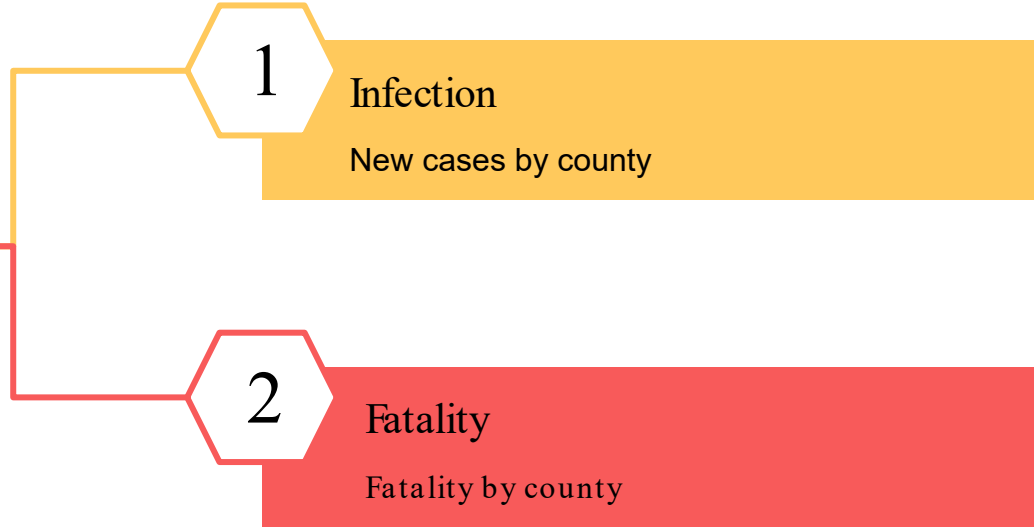
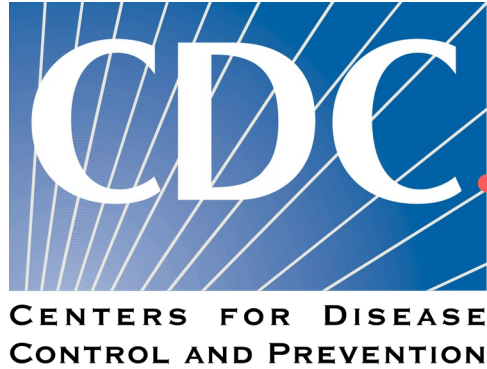
# Roadmap



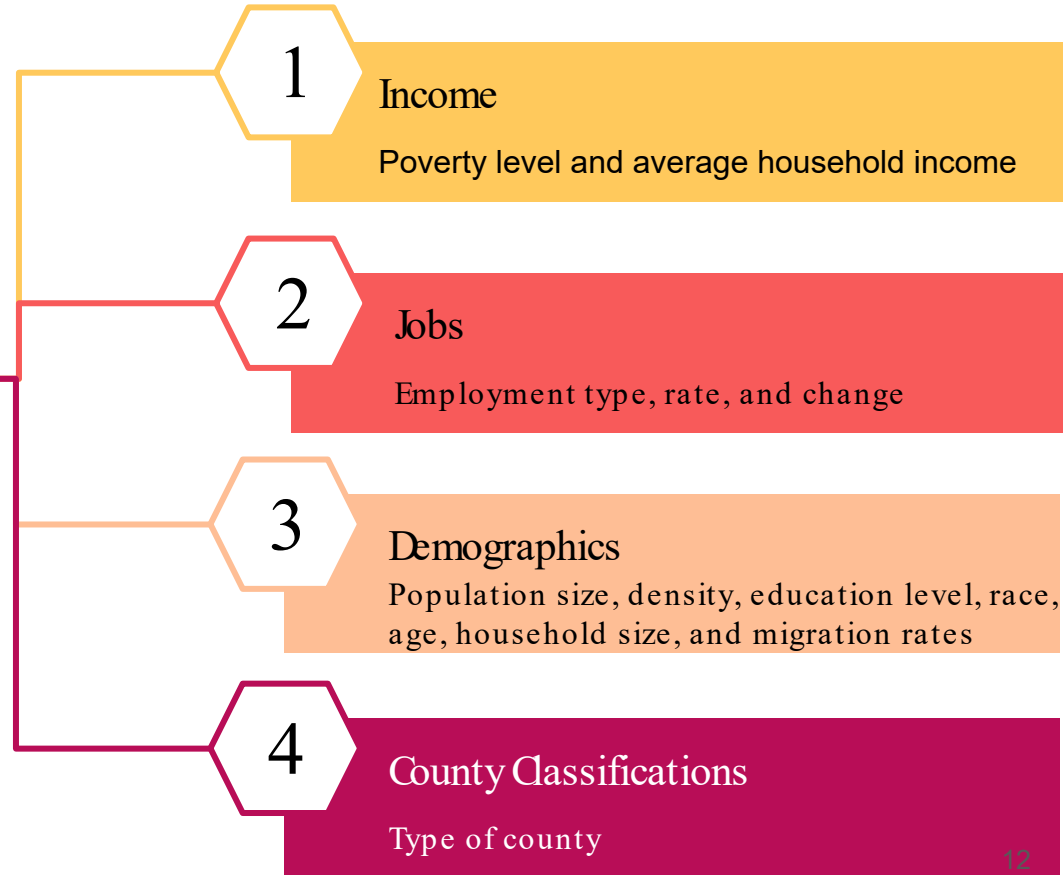
# Data



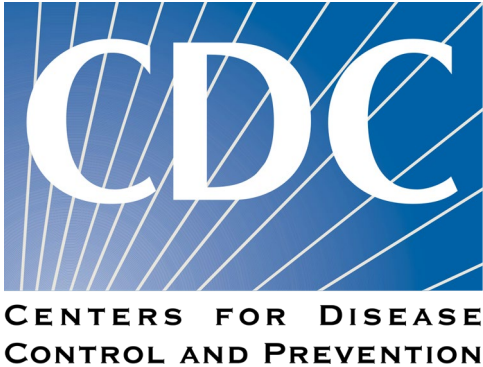
# Data



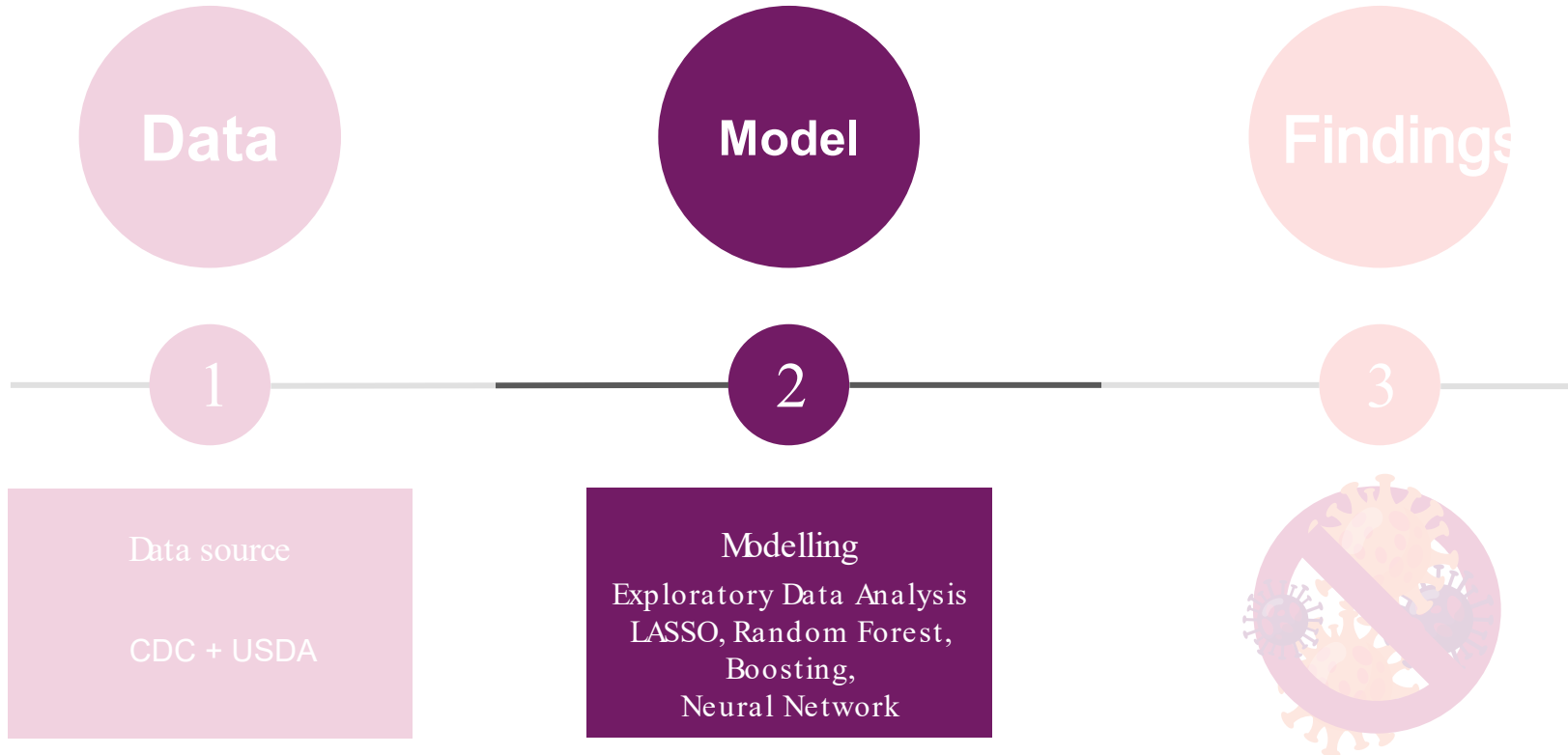
## Data



# Merging Data



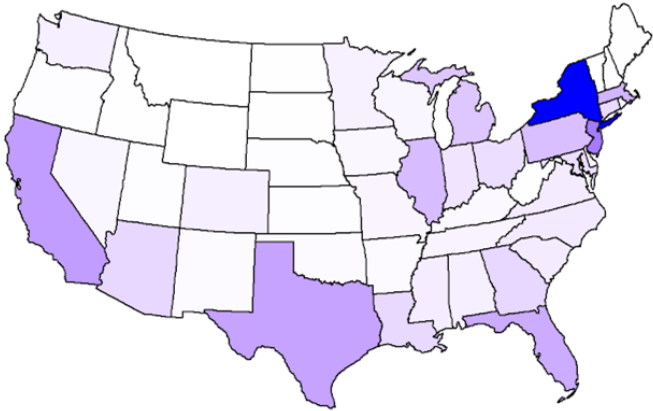
# Roadmap



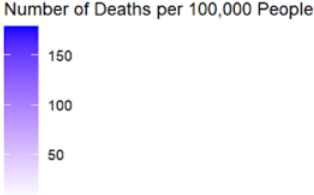
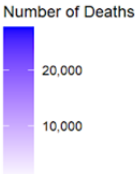
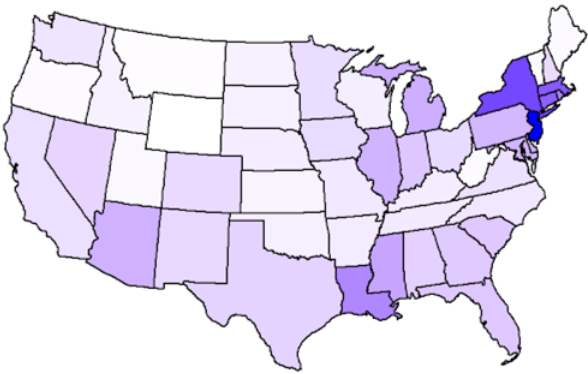


# Accumulative Deaths and Death Rate (8/ 19/ 2020)

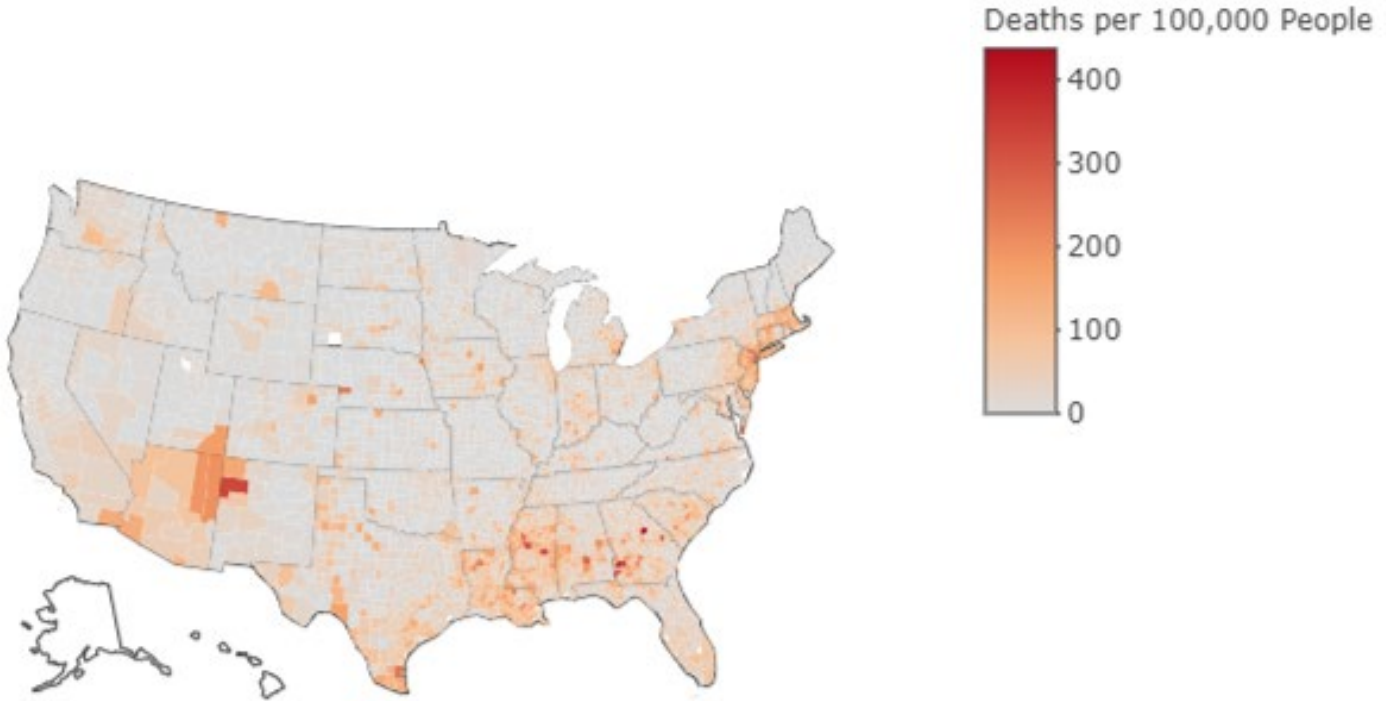
State Fatality Number  
Continental US States



State Fatality Rate  
Continental US States

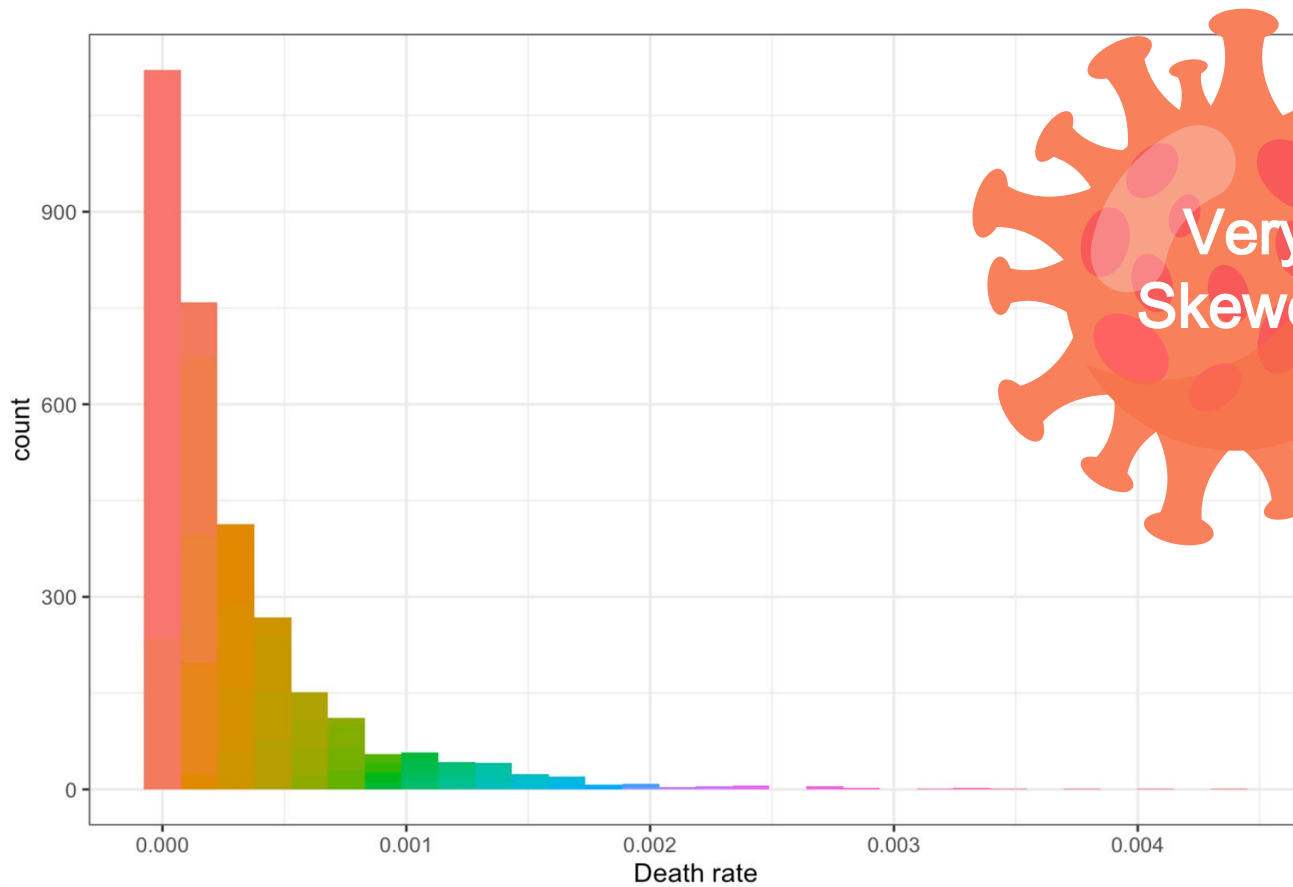


# Death rate by county (8/ 19/ 2020)

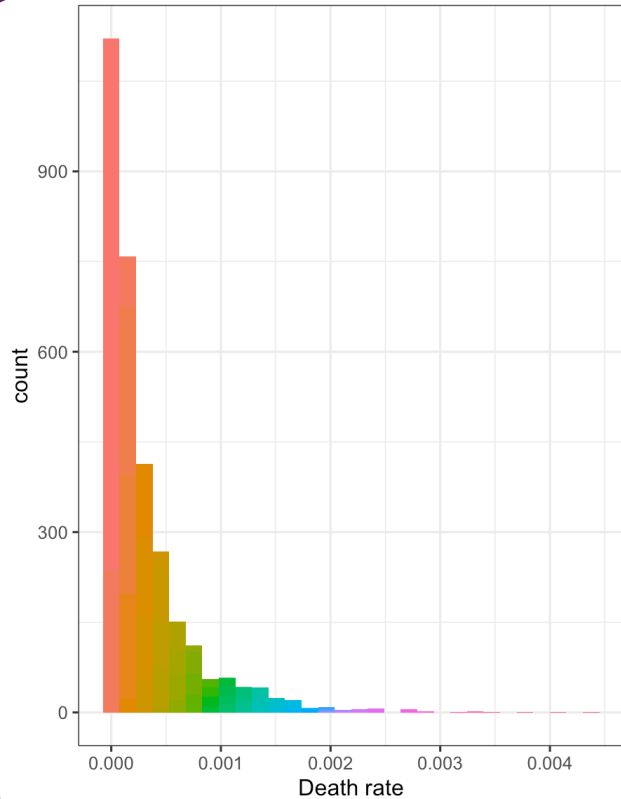




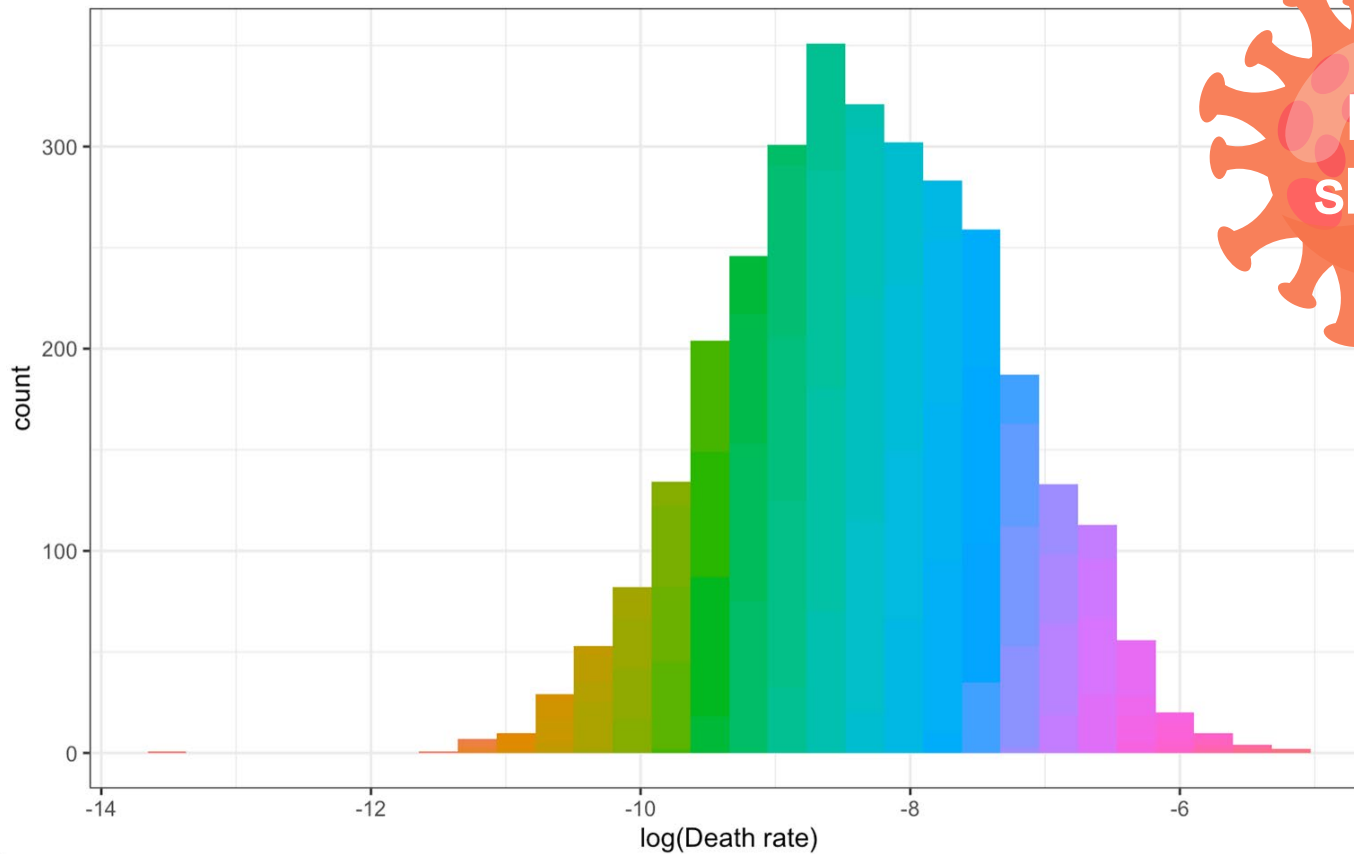
# Histogram of Death Rate



## Transformation



$$\log \left( \frac{\text{COVID-19 deaths} + 1}{\text{County population}} \right)$$

Histogram of  $\log(\text{Death Rate})$ 

Model

# LASSO Regression

Input

40+

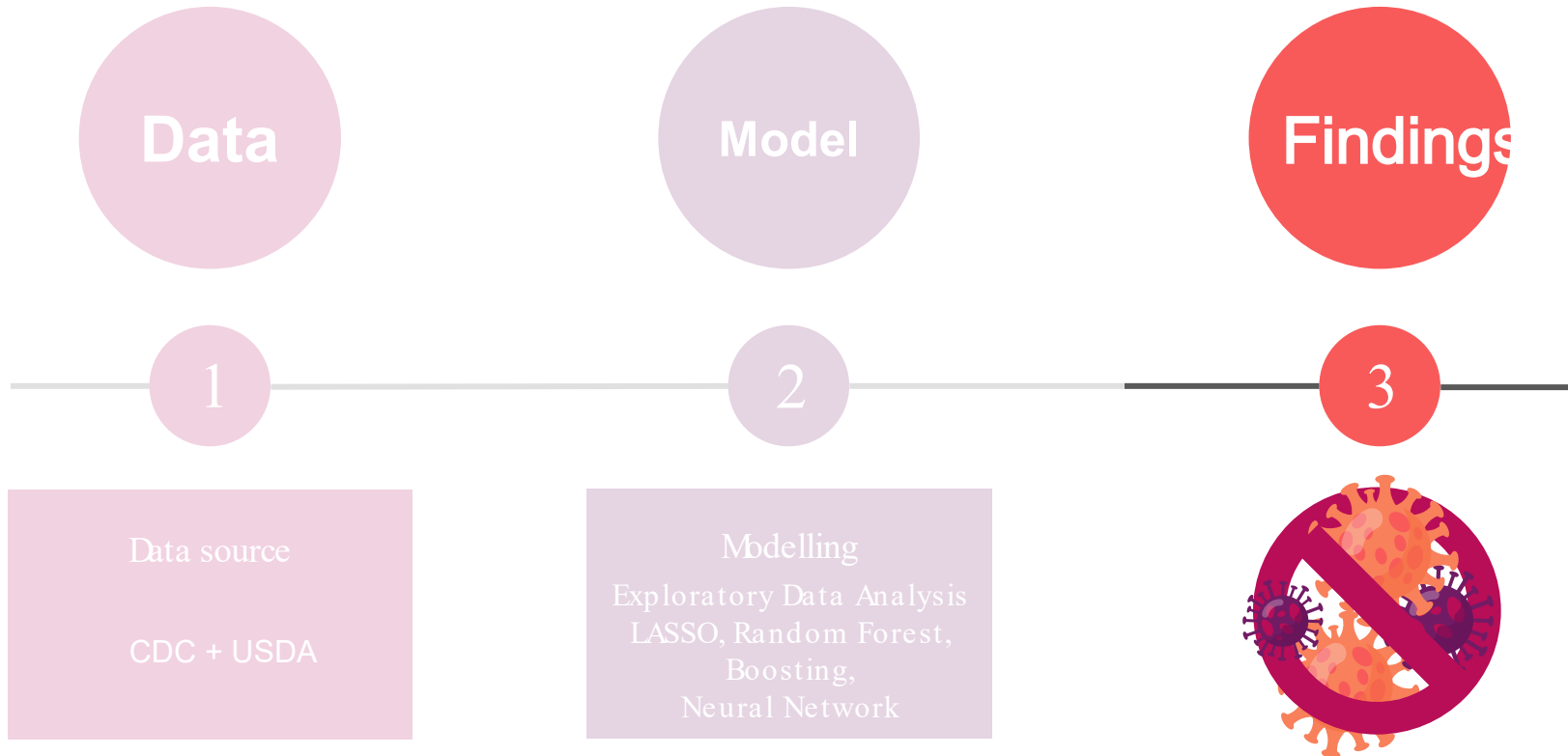
LASSO

Output

20

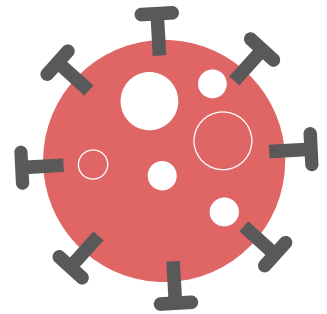


# Roadmap



Finding

# Worst States



NJ



CT



LA



MS



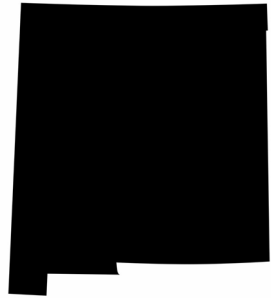
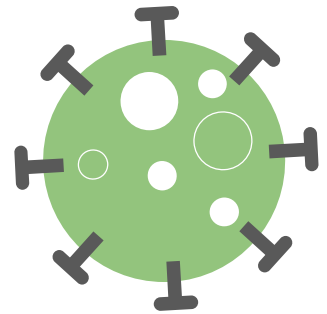
MA



DE



# Best States



**OR**

**CA**

**ME**

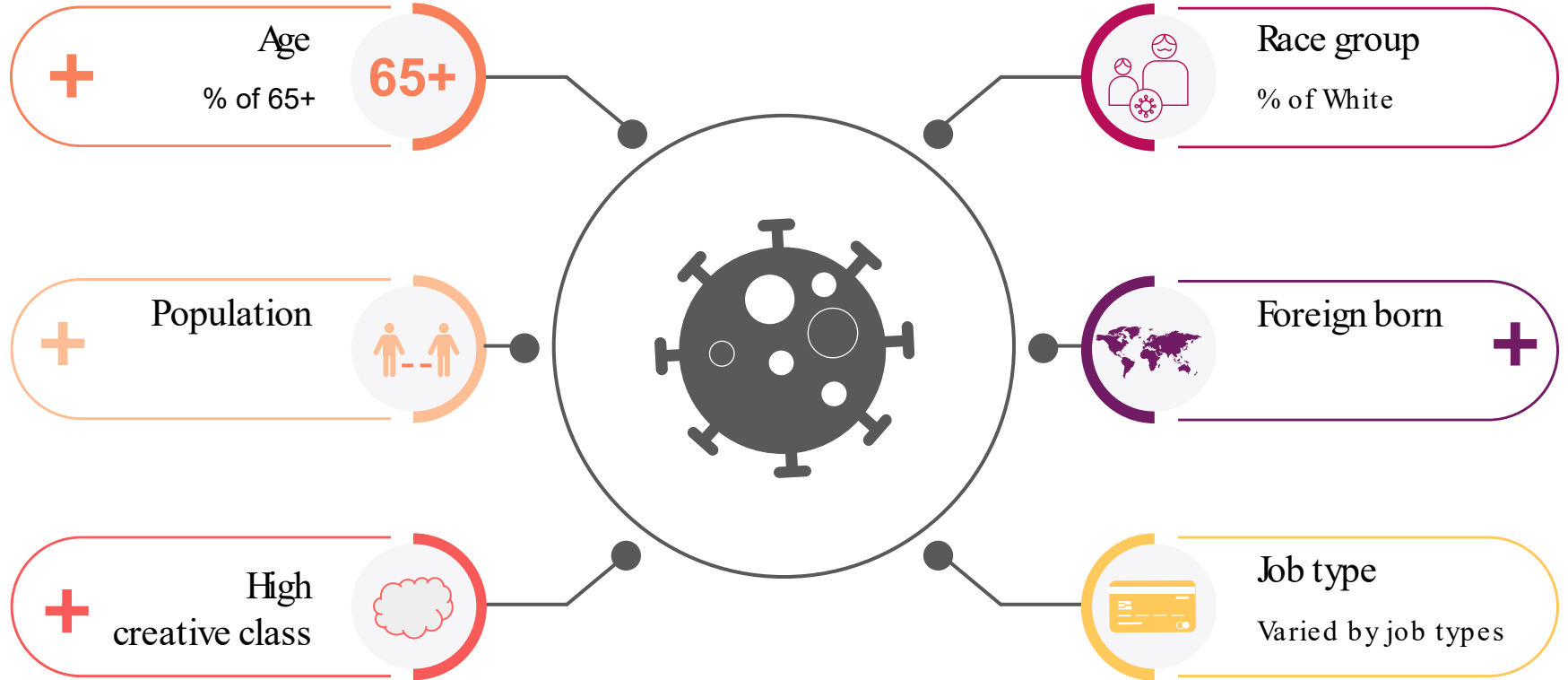
**WI**

**NM**

**WA**

Finding

# Covid Factors affecting death rate





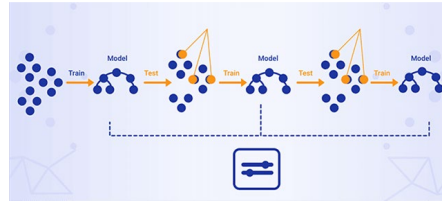
# Predictive Models

## Prediction Error



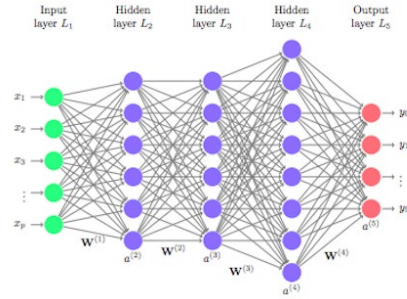
Random Forest

0.47



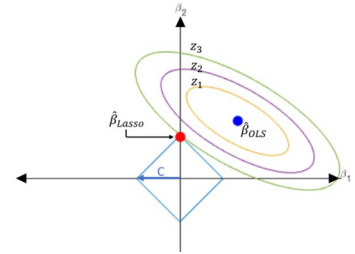
Boosting

0.50



Deep Learning

0.54



Relaxed LASSO

0.56

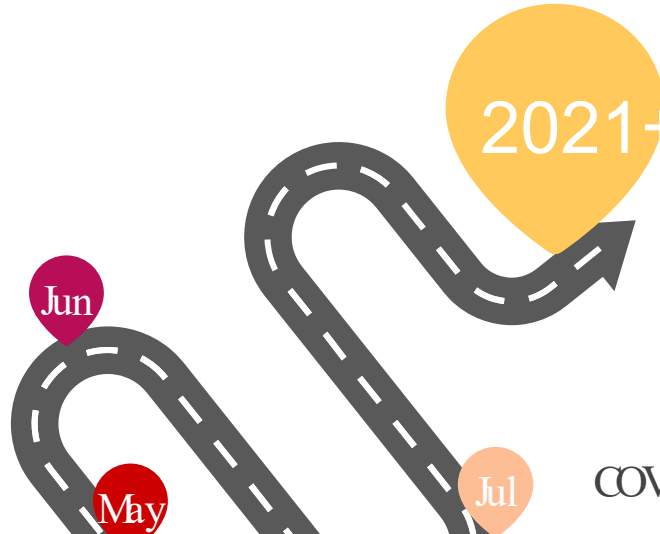
# My COVIDpath to Data Scientist

Thank you!

Acknowledgement:  
Professor Linda Zhao and Jeffrey Cai

Prof Zhao's  
Modern Data Mining Course

What can I do to help?



Data Scientist

Aim to be a data scientist  
Use data for social good

COVID+County study